

Visual Social Network Analytics for Relationship Discovery in the Enterprise

VAST2011-best paper

1. 摘要

- 一套系统
 - 利用企业内社交媒体信息, 挖掘/聚合/推理出社交图
 - 450,000人/75,000,000关系
- 实验细节
 - 结合覆盖12个月, 大范围, 1800人的显著用例的日志和访谈/面试数据
- 结合社会地位、佐证、特征, 用户可利用SaNDVis在企业环境中呈现已有关系、建立新的关系

2. 介绍

- 信息探索
 - 问题: 关心的数据太多
 - 目前方案: 用复杂的信息探索工具定位数据 (如网页搜索引擎), 这些工具对数据进行爬取、挖掘、排名
 - 特点: 工具是document-centric (包括索引和交互)
 - 关系探索
 - 问题: 不是所有的信息探索都关心文档
 - 寻找专家/合作伙伴, 建立关系
 - 呈现已有关系, 观察企业内信息流动、人员合作
 - 把people-centric任务认为是**关系探索**任务
 - 方案: 关系探索的核心是社交图, 从带社会属性的文档中抽取社交图
 - 从公司文档中抽取
 - 传统媒体
 - 文章
 - 专利
 - 组织图
 - 在线社交媒体
 - 博客
 - 书签
 - 社区
 - 与信息探索的不同
 - 需要理解他们是谁? 如何沟通? 什么联系?
 - 他们被搜索到的原因 - 佐证
 - 他们可以被过滤 - 特征
 - 文章贡献
 - 企业内的关系探索的端到端分析系统
 - 系统组件
 - SaND
 - 从数据源挖掘&聚合社交媒体 (数据)
 - SaNDGraph
 - 将数据组织入以人为中心的数据库, 支持社交图的快速查询
 - SaNDVis
 - 用户管理复杂多维信息的UI
 - 展示社交图
 - 高亮佐证并链接到相应文档
 - 文章结构
 - 相关工作
 - 社会匹配 Social Matching
 - 专家定位 Expertise Location
 - 社交网络可视化分析研究
 - 系统介绍
 - SaND
 - SaNDGraph
 - SaNDVis
 - 评估-用户使用情况
 - 日志
 - 访谈
 - 优缺点与总结
- ## 3. 相关工作
- ### 1. 社会匹配&专家定位
- 社会匹配的动机
 - dating

- pursuing shared interests
 - addressing community issues
 - solving technical problems
 - having a good conversation
- 专家定位系统
 - 输入：查询
 - 输出：可能的专家列表
 - 相关技术
 - 基于文档文本的用户分析
 - Expert Finder: 从java文档中区分新手和专家
 - 大型组织中专家定位的动机大多数为：回答技术问题&找人
- 专家定位系统+社交数据
 - ReferralWeb: 结合社交网络，用户可指定搜索主题和社交标准
 - 如：找和Joh Smith最多有两重关系的人
 - Expertise Recommender: 通过用户的组织关系和社交关系过滤专家搜索结果
 - 借鉴其中的关系图
 - 基于内容和社交图，提高专家识别的有效性
- 社交媒体新技术
 - Collabio: Facebook给好友贴标签
 - Farrell: 公司内给他人标签

2. 社交网络可视化分析研究

- 视图
 - node-link diagrams
 - matrix visualizations
 - rolledup PivotGraphs
 - multiple coordinated view
 - tabular view
- 个人关系网络
 - PersonalMap
 - ContactMap
 - Soylent
 - Post History
 - Social
 - Network Fragments
 - Vizster
- 特定主题的社交网络
 - SmallBlue

4. SOCIAL NETWORKS & DISCOVERY (SaND)

1. 关系聚合

- SaND是关系聚合分析系统
 - 分析“人/文档/标签”之间的关系
 - 相对于之前的“人”之间的关系（SONAR），增加了社交和文本的交叉分析
- 关系模型图



■ Figure 1. Direct Relationships Modeled in SaND

- 关系带权重
 - 提供查询API
- 2. 数据源
 - 数据是公开的

- UI界面

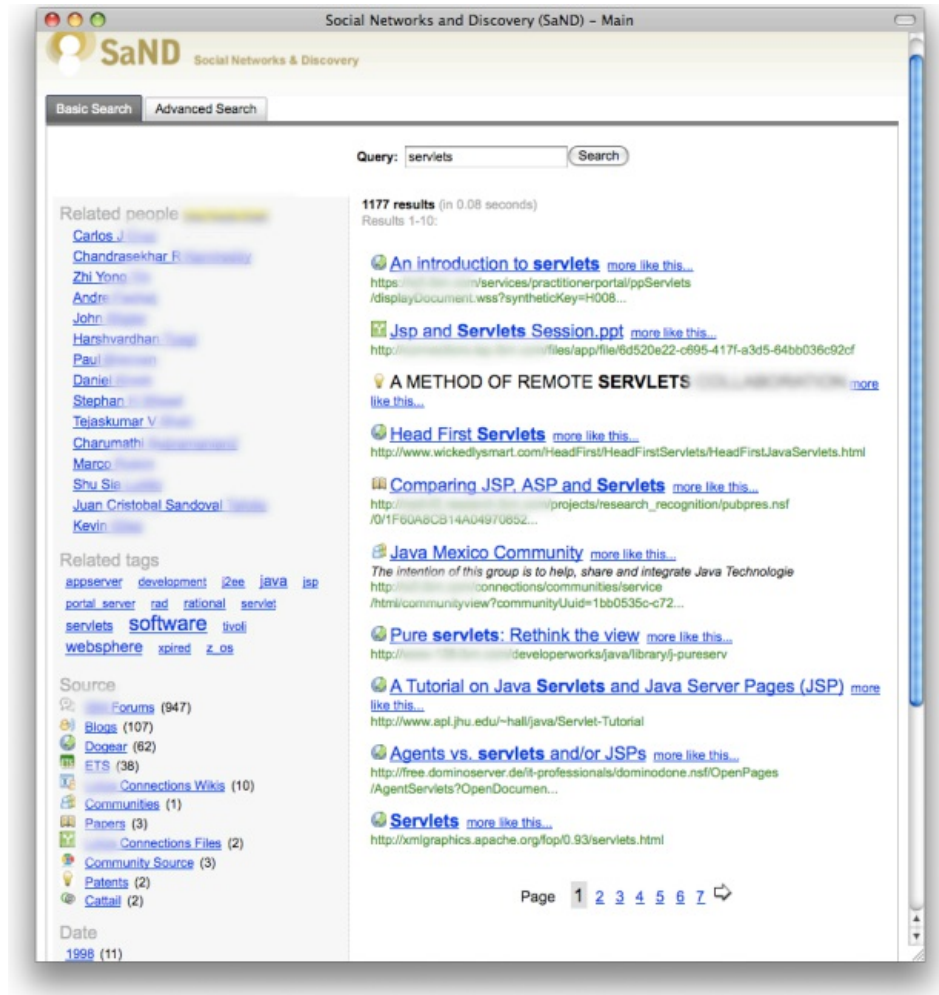


Figure 2. SaND's Social Search UI.

- 实际数据源

- A blogging system with 16,300 blogs, 144,200 entries, 121,750 comments, 70,000 overall users and 357,000 tags
- A wiki system with 6,150 public wikis, 13,000 editors and 24,450 tags
- A social bookmarking system with 1.1M bookmarks by 68,000 users with 3M tags
- A file sharing system with 46,700 public files used by 31,800 users and tagged with 86,000 tags
- A community system with 9,400 online communities, each including resources such as feeds and forums, with an overall of 226,000 members and 32,500 tags
- An organizational chart including nearly 450,000 employees
- Two enterprise SNSs that allow users to reciprocally connect to each other, with an overall of 250,000 connections between 99,000 users
- A patent database with 132,000 patents authored by 31,500 users
- A publication database that includes 28,950 papers authored by 3,200 users
- A projects wiki that includes 1,980 projects with 1,260 members and 2,450 tags
- An open source project system with 1,860 projects and 11,850 total members
- A forum system with 2,590 forums, 466,300 threads and 53,000 users
- A people tagging application that allows users to tag each other, with 9,300 users who tagged 50,000 other individuals with 160,000 public tags

3. 从文档到人：支持关系探索

- 以文档为中心的UI
 - 有利于文本信息的发现
 - 基于文档建立索引
- 以人为中心的UI
 - 有利于关系的发现
 - 基于人建立索引
 - SaNDGraph数据库
 - 支持实时查询450,000人/73,000,000关系
 - 抽取“人-人”关系
 1. 关系分为6类
 - organizational

- friending
- tagging
- commenting
- coauthorship
- co-membership

2. 辨别出可以反映熟悉度/相似度的关系

	Familiarity	Similarity
Organizational Chart	Being manager or employee Sharing a 1st- or 2nd-level manager	
Friending	Being friends on an SNS	Having a common friend on an SNS
Tagging	Tagging a person Being tagged by a person	Co-tagging the same document Co-tagging the same person Co-usage of the same tag Being tagged by the same person Being tagged with the same tag
Commenting	Commenting to a person's document (blog entry or file) Comment by a person to own document (blog entry or file)	Co-commenting on the same document (blog entry or file)
Co-authorship	Co-authorship of a patent, paper, file, or wiki	Being shared a file by the same person Corresponding on the same forum thread
Co-membership	Co-membership in a small community (≤ 20 total members) or project	Co-membership in a large community (> 20 total members)

Table 1. A classification all of the people-to-people relationships indexed by SaNDGraph. Familiarity relationships indicate the connected people probably know each other. Similarity relationships

- indicate the connected people behave similarly across social media.

5. SaNDVIS : 关系探索系统

- 利用SaNDGraph社交图查询的以人为中心的可视分析UI

- 输入：文本查询
- 输出：
 - 关系图视图：匹配的人
 - 佐证视图：佐证总览和标签
 - 特征视图

1. 关系图视图

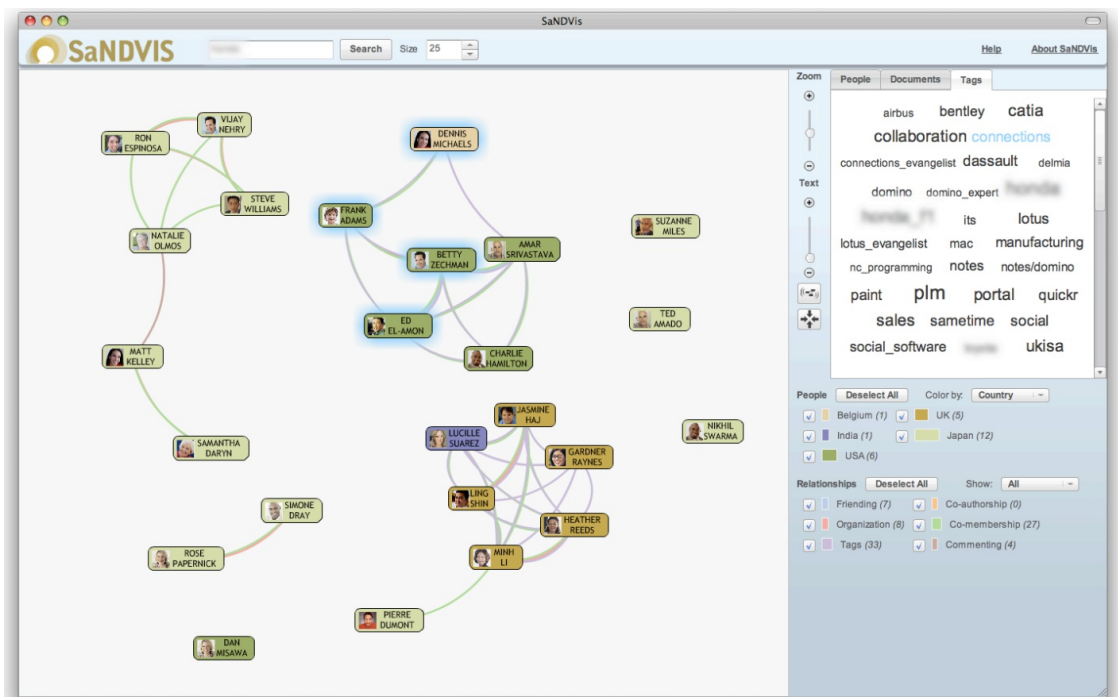


Figure 3. The SaNDVis UI contains three views relevant to relationship discovery task. A) On the left, a social graph view allows users to interpret the social position of the results of a topic query. B) On the top right, an evidence view allows users to examine the documents, tags, and people associated with the query. C) On the bottom right, a facet view allows users to get an overview of the categories with sparklines and allows users to filter out any irrelevant categories.

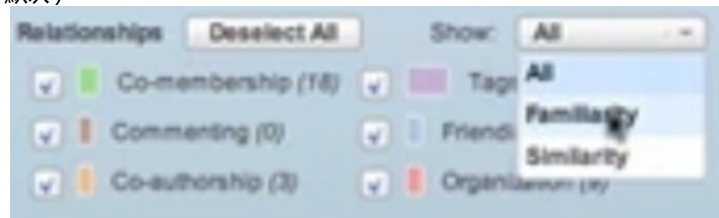
- 人
 - 头像
 - 名字
 - 高亮社会地位（没看出来）
 - 用于估计实用的关系
 - 颜色：编码人的分类（国家或部门，参见5.3）
 - 数量：默认最相关的25人，可改变
- 关系
 - 颜色：编码关系种类
 - 宽度：编码权重
 - 相同两人间的关系整合为一条彩虹边
- 算法
 - 加强版力引导算法
 - stress majorization algorithm 应力控制算法
 - 最小化节点重叠和边交叉

2. 佐证视图

- 标签



- 根据部门分类
- 关系
 - 参见Table 1的6大类关系
 - 熟悉度
 - 相似度
 - All (默认)



- sparkline & checkbox
- 社交图联动
 - 人
 - 过滤后重新布局，自动缩放
 - 关系
 - 不重新布局

6. 评估

- 跨国IT公司/1790人/50个国家/15674关系/2010.3~2011.3/用户提供唯一id
- 1. 关系探索查询分类

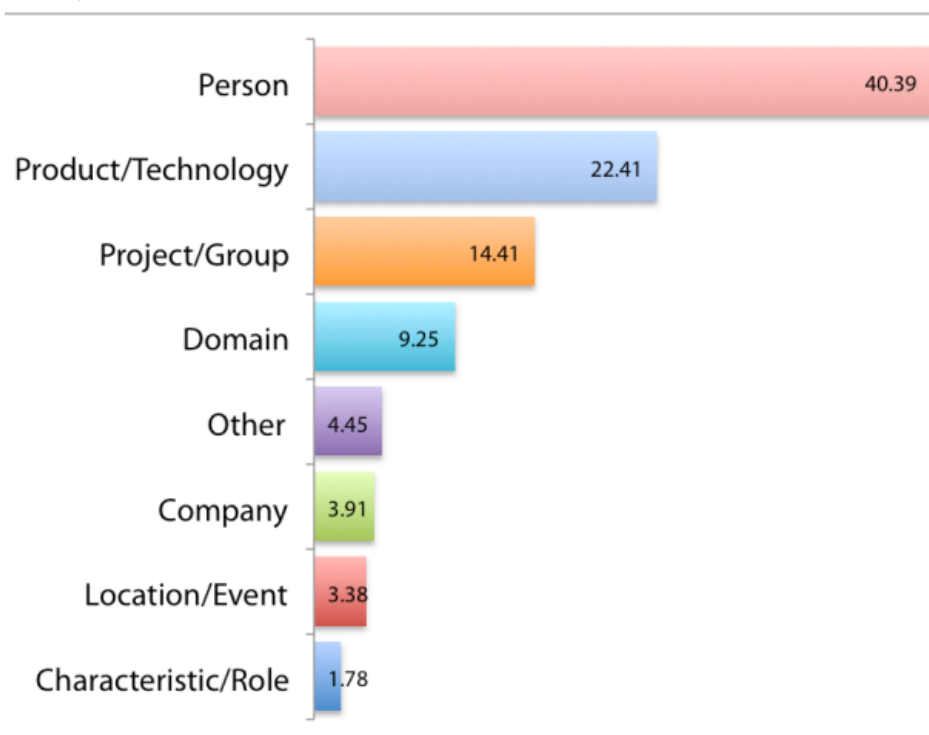


Figure 4. Percentages of the different categories of user queries using SaNDVis.

- 特定某人的关系
- 与某技术或产品相关的关系
- 内部项目与小组
- 专业术语
- 公司术语
- 地址/事件
- (人的) 特点/角色

7. 用例

1. 专家定位&团队组建
2. 团队协作和反馈
3. 个人社交反馈
 - 以人为核心方法的需求

- 有时候找的就是人
- 比社交网站透明

8. 讨论&限制

1. 仅限于活跃的社交媒体信息的探索
2. 用户不想总是用关键词查询，希望可以search for explicit group
3. 难以扩展到公司外，实例匹配，数据规模等问题
4. UI是新颖的，而不是自行发展的

9. 总结

- 一个端到端的支持企业内关系探索的可视分析系统
 - 功能：从文档中挖掘/聚合/整合社交图
 - 用途：
 - 专家定位
 - 团队组建
 - 个人社交反馈
 - 亮点：
 - 传统工具不能很好的支持以上任务
 - 结合社会地位、佐证、特征，帮助用户分析关系
 - 有一个不错的实验结果

10. 引用